Dirac's quantum jump

Atkinson, D

Published in:
Physics Education

Publication date:
1999

Professor D. Atkinson
Institute for Theoretical Physics
University of Groningen
Nijenborgh 4
NL-9747 AG Groningen
The Netherlands
E-mail: atkinson@phys.rug.nl

Tel: - 31-50-3634953 Fax: - 31-50-3634947

Dirac's Quantum Jump

D. Atkinson

Lecture given at Pune University

January 1999

Foreword

This minicourse on quantum mechanics is intended for students who have already been rather well exposed to the subject at an elementary level. It is assumed that they have surmounted the first conceptual hurdles and also have struggled with the Schrödinger equation in one dimension.

It is always possible to criticize the level of mathematical sophistication and, indeed, correctness of an exposition of quantum mechanics. So be it. Sufficient unto the course is the rigour thereof!

Chapter 1

Hamilton's Classical Mechanics

In this chapter we explain briefly Hamilton's formulation of classical mechanics, and Dirac's inductive transition to quantum mechanics.

1.1 Hamilton's Equations

Consider a system of N particles in a conservative field. We write the kinetic energy in the form

$$T = \tfrac{1}{2} m_i \dot{x}_i \dot{x}_i, \tag{1.1}$$

where we sum over the Cartesian coordinates of all N particles, with the convention that $m_{3i-2} = m_{3i-1} = m_{3i}$ is the mass of the i'th particle, the x, y, z coordinates of which are x_{3i-2}, x_{3i-1} and x_{3i} respectively. The potential energy may include particle-particle interactions, so long as these are conservative:

$$V = \sum_{n=1}^{N} V_n + \sum_{n=1}^{N} \sum_{p=1}^{n-1} V_{np}. \tag{1.2}$$

We shall now perform a (possibly explicitly time-dependent) transformation of the $3N$ Cartesian coordinates, x_i, to a set of $3N$ independent coordinates, q_n. A simple example would be the polar coordinates of the N particles, but much more exotic possibilities exist. The purpose is eventually to obtain coordinate-independent equations of motion.

Define the Lagrangian of the system by $L = T - V$. This is quite different from the total energy, $E = T + V$. Since V does not depend explicitly on $\dot{x}_i$, neither does it depend explicitly on $\dot{q}_n$, which obviously implies that the partial $\dot{q}_n$-derivative of T is equal to that of L. After some work we find that Newton's equations of motion can be rewritten as follows:

$$\frac{d}{dt}\left[\frac{\partial L}{\partial \dot{q}_n}\right] - \frac{\partial L}{\partial q_n} = 0, \tag{1.3}$$

which is the famous Euler-Lagrange equation.

In terms of Cartesian components, the Lagrangian of our system of N particles, interacting with one another, and with an external, conservative field, can be written

$$L(x, \dot{x}, t) = \tfrac{1}{2} m_i \dot{x}_i \dot{x}_i - V(x, t), \tag{1.4}$$

where V may depend explicitly on t (for example, a time-dependent external potential), but not on $\dot{x}$. Then clearly

$$\frac{\partial L}{\partial \dot{x}_i} = m_{(i)} \dot{x}_{(i)}, \tag{1.5}$$

where there is no summation over the bracketed index, (i). We recognize the right-hand side of Eq.(1.5) as a component of the linear momentum, and we now define the generalized, or canonical momentum, by

$$p_n = \frac{\partial L}{\partial \dot{q}_n}. \tag{1.6}$$

It is important to realize that the canonical momenta are not necessarily momenta in the Cartesian sense: for example, in terms of spherical polar coordinates, the "momentum" conjugate to the angle, θ, is the angular momentum.

The Hamiltonian is defined by

$$H = p_n \dot{q}_n - L, \tag{1.7}$$

with summation over n, and it can be shown that

$$\frac{dH}{dt} + \frac{\partial L}{\partial t} = 0. \tag{1.8}$$

If L does not depend explicitly on the time, the second term above is absent, and hence the Hamiltonian is constant in time. If, moreover, the potential energy is conservative, Eq.(1.5) holds, so that

$$H = m_i \dot{x}_i \dot{x}_i - L = 2T - [T - V] = T + V, \tag{1.9}$$

which means that the Hamiltonian is equal to the total energy, which is time-independent. Note that two conditions are necessary for these conclusions to be true: (1) L must not explicitly depend on the time, and (2) V must not explicitly depend on the $\dot{q}$.

From the definition, Eq.(1.7), we deduce

$$dH = \dot{q}_n dp_n + p_n d\dot{q}_n - \frac{\partial L}{\partial q_n} dq_n - \frac{\partial L}{\partial \dot{q}_n} d\dot{q}_n - \frac{\partial L}{\partial t} dt. \tag{1.10}$$

The second and the fourth terms above cancel. Moreover the Euler-Lagrange equation implies that

$$\frac{\partial L}{\partial q_n} = \dot{p}_n, \tag{1.11}$$

so that

$$dH = \dot{q}_n dp_n - \dot{p}_n dq_n - \frac{\partial L}{\partial t} dt. \tag{1.12}$$

The form of this increment suggests that we consider H to be a function of q_n, p_n and t (and not independently of $\dot{q}_n$). Then the following partial derivatives can be read off from Eq.(1.12):

$$\dot{q}_n = \frac{\partial H}{\partial p_n} \tag{1.13}$$

$$\dot{p}_n = -\frac{\partial H}{\partial q_n} \tag{1.14}$$

$$\frac{\partial L}{\partial t} = -\frac{\partial H}{\partial t} = -\frac{dH}{dt}. \tag{1.15}$$

These are the Hamilton equations, which constitute an alternative to the Euler-Lagrange system.

One of the advantages of Hamilton's approach is that one can readily define more general transformations of the variables. The set $\{q_n, p_n\}$ may be replaced by another set $\{Q_n, P_n\}$ in which the Q's and the P's can be functions of both the q's and the p's, and possibly of t. If the Hamilton equations, Eq.(1.13)–Eq.(1.15), remain valid in terms of the new variables (i.e. the transformation is such that these equations, with q_n and p_n replaced respectively by Q_n and P_n, are true), then we speak of a canonical transformation. The new variables are just as acceptable as the old ones. P_n is called the momentum canonically conjugate to Q_n.

1.2 Poisson Brackets

Let (q_j, p_j) be a set of canonical coordinates and their conjugate momenta. The Poisson bracket of two functions of them is defined by

$$[B, C] = \frac{\partial B}{\partial q_j}\frac{\partial C}{\partial p_j} - \frac{\partial B}{\partial p_j}\frac{\partial C}{\partial q_j}, \tag{1.16}$$

with implicit summation over j. Clearly

$$[B, C] = -[C, B]. \tag{1.17}$$

Let ξ_k be one of the q_j or one of the p_j. Then

$$\begin{aligned} \frac{\partial}{\partial \xi_k}[B, C] &= \frac{\partial^2 B}{\partial q_j \partial \xi_k}\frac{\partial C}{\partial p_j} - \frac{\partial^2 B}{\partial p_j \partial \xi_k}\frac{\partial C}{\partial q_j} + \frac{\partial B}{\partial q_j}\frac{\partial^2 C}{\partial p_j \partial p_j \partial \xi_k} - \frac{\partial B}{\partial p_j}\frac{\partial^2 C}{\partial q_j \partial \xi_k} \\ &= \left[\frac{\partial B}{\partial \xi_k}, C\right] + \left[B, \frac{\partial C}{\partial \xi_k}\right]. \end{aligned} \tag{1.18}$$

We may write

$$[B, C] = D_B C, \tag{1.19}$$

where D_B is the linear differential operator

$$D_B = \frac{\partial B}{\partial q_j}\frac{\partial}{\partial p_j} - \frac{\partial B}{\partial p_j}\frac{\partial}{\partial q_j} \equiv \sum_k \beta_k \frac{\partial}{\partial \xi_k}. \tag{1.20}$$

Here ξ_k runs the q's *and* the p's. The important point is that the functions β_k depend on B, but not on C. Similarly,

$$[A, C] = D_A C \tag{1.21}$$

where

$$D_A = \sum_k \alpha_k \frac{\partial}{\partial \xi_k}, \tag{1.22}$$

with α_k independent of C.

Consider

$$\begin{aligned} [A,[B,C]] &= D_A D_B C = \sum_{i,j} \alpha_i \frac{\partial}{\partial \xi_i} \beta_j \frac{\partial}{\partial \xi_j} C \\ &= \sum_{i,j} \left(\alpha_i \frac{\partial \beta_j}{\partial \xi_i} \frac{\partial}{\partial \xi_j} + \alpha_i \beta_j \frac{\partial^2}{\partial \xi_i \partial \xi_j} \right) C. \end{aligned} \tag{1.23}$$

Clearly

$$\begin{aligned} [A,[B,C]] - [B,[A,C]] &= (D_A D_B - D_B D_A) C \\ &= \sum_j \left\{ \sum_i \left(\alpha_i \frac{\partial \beta_j}{\partial \xi_i} - \beta_i \frac{\partial \alpha_j}{\partial \xi_i} \right) \right\} \frac{\partial}{\partial \xi_j} C, \end{aligned} \tag{1.24}$$

since the second-order derivatives cancel. This can be rewritten in terms of the q_i and the p_i in the form

$$[A,[B,C]] - [B,[A,C]] = \left(\gamma_k \frac{\partial}{\partial q_k} + \delta_k \frac{\partial}{\partial p_k} \right) C, \tag{1.25}$$

once more with implicit summation.

Since we have proved that the functions γ_k and δ_k depend only on A and B, but not on C, we can evaluate them by substituting simple functions for C on both sides of Eq.(1.25). With $C = q_i$, we find

$$\gamma_i = -\left[A, \frac{\partial B}{\partial p_i} \right] + \left[B, \frac{\partial A}{\partial p_i} \right] = -\frac{\partial}{\partial p_i} [A, B], \tag{1.26}$$

where we have used Eq.(1.17) and Eq.(1.18). Similarly, with $C = p_i$, we have

$$\delta_i = \left[A, \frac{\partial B}{\partial q_i} \right] - \left[B, \frac{\partial A}{\partial q_i} \right] = \frac{\partial}{\partial q_i} [A, B]. \tag{1.27}$$

On substituting Eq.(1.26) and Eq.(1.27) into Eq.(1.25), we now have, for general C,

$$\begin{aligned} [A,[B,C]] &- [B,[A,C]] \\ &= -\left\{ \frac{\partial}{\partial p_k} [A, B] \right\} \frac{\partial C}{\partial q_k} + \left\{ \frac{\partial}{\partial q_k} [A, B] \right\} \frac{\partial C}{\partial p_k} \\ &= -[C,[A,B]]. \end{aligned} \tag{1.28}$$

Finally, with use of Eq.(1.17), we can write the **Jacobi Identity for Classical Poisson Brackets** in the form

$$[A,[B,C]] + [B,[C,A]] + [C,[A,B]] = 0. \tag{1.29}$$

1.3 Dirac's Quantum Jump

The time-derivative of an arbitrary function, F, of the canonical variables (q_n, p_n), and possibly explicitly of the time, can be written

$$\begin{aligned} \frac{dF}{dt} &= \frac{\partial F}{\partial q_n}\dot{q}_n + \frac{\partial F}{\partial p_n}\dot{p}_n + \frac{\partial F}{\partial t} \\ &= \frac{\partial F}{\partial q_n}\frac{\partial H}{\partial p_n} - \frac{\partial F}{\partial p_n}\frac{\partial H}{\partial q_n} + \frac{\partial F}{\partial t} \\ &= [F, H] + \frac{\partial F}{\partial t}\,. \end{aligned} \tag{1.30}$$

The total time-derivative of a function that does not depend *explicitly* on the time is given by the Poisson bracket of that function with the Hamiltonian: we say that the Hamiltonian *generates* time-translations. In particular,

$$\dot{q}_n = [q_n, H], \tag{1.31}$$

$$\dot{p}_n = [p_n, H]. \tag{1.32}$$

These equations are impressively symmetrical between the q's and the p's, and they replace the first two of the Hamilton equations, (1.13) and (1.14). It is easy to see that

$$[q_k, p_l] = \delta_{kl}, \tag{1.33}$$

and Dirac's recipe for the intuitive jump from classical to quantum mechanics is to represent dynamical variables by linear operators on a Hilbert space, and Poisson brackets by commutators (multiplied by $1/(i\hbar)$). With this interpretation, the dynamics Eq.(1.31)-Eq.(1.32), as well as the "quantization condition" Eq.(1.33), have the same form in classical and in quantum mechanics.

"In setting up this form of the equations, Hamilton was influenced only by conditions of mathematical beauty. He might have said: 'It is very nice to write the equations in this way, but ... you could ... continue to use the equations in the form they were originally given by Newton.' But Hamilton seemed to have some remarkable insight into what was important—one of the most remarkable insights, I suppose, that a mathematician has ever had. He found a form of writing the equations of mechanics whose importance would be realized only after a hundred years.

It is more important for our equations to be beautiful than to have them fit experiment.

We had these equations involving these noncommuting quantities, but to begin with, we had no general interpretation for the equations. That was really a remarkable situation to have in a physical theory. (In any physical theory one usually knows just what one's equations mean before one sets them up.)

De Broglie was led to this idea of connecting waves with particles, by mathematical beauty.

It was found that this equation[1] gave the particle a spin of half a quantum. And also gave it a

[1] The Dirac equation

magnetic moment. It gave just the properties that one needed for an electron. That was really an unexpected bonus for me, completely unexpected.

Now, with quantum mechanics, we cannot exclude transitions from positive energy states to negative energy states, and that means that we cannot exclude the negative energy states from our theory. ... We can get a departure from the vacuum in two ways: one way is to bring an electron into a positive energy state; the other way is to have a 'hole' in the distribution of negative energy states ... the 'holes' appear as a new kind of particle having a positive charge. What is the mass of these new particles? Well, when I first thought of this idea, it occurred to me that the mass would have to be the same as that of the electron because of the symmetry. But I did not dare to put forward that idea, because it seemed to me that if this new kind of particle (having the same mass as the electron and an opposite charge) existed, it would certainly have been discovered by the experimenters. ... I think that Weyl was the first to make the very definite statement that mathematical symmetry demanded that these 'holes' should be particles with the same mass as the mass of the electron. ... and the question arises: 'Why had experimenters never observed them?' I think the only answer to that question is that they were prejudiced against new particles. It was assumed that there were only two basic particles in Nature: the electron and the proton. ... They had never observed positrons, because they really turned a blind eye to them when they had evidence for them.

A great deal of my work is just playing with equations and seeing what they give."[2]

[2] P.A.M. Dirac, 'Directions in Physics', Wiley (1978)

Chapter 2

Canonical Quantization

2.1 Commutators

Consider the motion of a particle in one dimension. In quantum mechanics, the position of the particle is represented by a hermitian operator, $\mathbf{q}$. Let $|x\rangle$ and $|y\rangle$ be two eigenvectors of $\mathbf{q}$ corresponding to eigenvalues x and y, i.e.

$$\mathbf{q}|x\rangle = x|x\rangle; \qquad \mathbf{q}|y\rangle = y|y\rangle.$$

$$x^*\langle x|y\rangle = \langle y|\mathbf{q}|x\rangle^* = \langle x|\mathbf{q}|y\rangle = y\langle x|y\rangle.$$

Setting $x = y$, we see that $x^* = x$, since $\langle x|x\rangle \neq 0$; thus the eigenvalues are real. When $x \neq y$, we have

$$(x - y)\langle x|y\rangle = 0, \tag{2.1}$$

which implies $\langle x|y\rangle = 0$, i.e. eigenvectors corresponding to different eigenvalues are mutually orthogonal.

Let us solve Eq.(2.1) as an equation for x, at a fixed value of y. In a space of continuous functions, the only possibility for $\langle x|y\rangle$ would be zero, since it is zero for all $x \neq y$; but this would mean $\langle x|x\rangle = 0$ and that implies that $|x\rangle$ is the null vector, which we have excluded. In the space of tempered distributions, a solution of Eq.(2.1) is

$$\langle x|y\rangle = \delta(x - y), \tag{2.2}$$

or indeed any multiple of it. The above amounts to a choice of (continuum) normalization.

Proof that Eq.(2.2) satisfies Eq.(2.1)

Let $g(x)$ be a continuous test-function. Then

$$\int_{-\infty}^{\infty} dx\, g(x)(x-y)\delta(x-y) = [(x-y)g(x)]_{x=y} = 0.$$

$\triangle$

Let $\mathbf{p}$ be the operator representing the momentum of the particle. In classical mechanics, we have that the Poisson bracket of q and p is

$$[q,p]_{PB} = \frac{\partial q}{\partial q}\frac{\partial p}{\partial p} - \frac{\partial q}{\partial p}\frac{\partial p}{\partial q} = 1.$$

In quantum mechanics we follow Dirac in maintaining this relation, but reinterpreting the Poisson bracket as the commutator of the corresponding operators $\mathbf{q}$ and $\mathbf{p}$, multiplied by a universal constant, say κ:

$$\kappa(\mathbf{qp} - \mathbf{pq}) = 1. \tag{2.3}$$

The important point is that, like the Poisson brackets in classical mechanics, commutators satisfy the following algebraical identities:

$$[\mathbf{q},\mathbf{p}] = -[\mathbf{p},\mathbf{q}]$$

$$[\mathbf{r},[\mathbf{q},\mathbf{p}]] + [\mathbf{q},[\mathbf{p},\mathbf{r}]] + [\mathbf{p},[\mathbf{r},\mathbf{q}]] = 0 \tag{2.4}$$

Proof of Eq.(2.4)

$$\begin{aligned}
[\mathbf{q},\mathbf{p}] &= \mathbf{qp} - \mathbf{pq} = -[\mathbf{p},\mathbf{q}] \\
[\mathbf{r},[\mathbf{q},\mathbf{p}]] &= \mathbf{rqp} - \mathbf{rpq} - \mathbf{qpr} + \mathbf{pqr} \\
[\mathbf{q},[\mathbf{p},\mathbf{r}]] &= \mathbf{qpr} - \mathbf{qrp} - \mathbf{prq} + \mathbf{rpq} \\
[\mathbf{p},[\mathbf{r},\mathbf{q}]] &= \mathbf{prq} - \mathbf{pqr} - \mathbf{rqp} + \mathbf{qrp}\,.
\end{aligned} \tag{2.5}$$

$\triangle$

From Eq.(2.3), we see that

$$\langle x|y\rangle = \kappa\langle x|\mathbf{qp} - \mathbf{pq}|y\rangle = \kappa(x-y)\langle x|\mathbf{p}|y\rangle. \tag{2.6}$$

We require $\mathbf{p}$ to be hermitian, thus ensuring that its eigenvalues are real. Then, by taking the complex conjugate of both sides of the above equation, we obtain

$$\langle y|x\rangle = \kappa^*(x-y)\langle y|\mathbf{p}|x\rangle. \tag{2.7}$$

In Eq.(2.7) rename x and y as, respectively, y and x:

$$\langle x|y\rangle = \kappa^*(y-x)\langle x|\mathbf{p}|y\rangle. \tag{2.8}$$

On comparing Eq.(2.6) and Eq.(2.8), we see that $\kappa = -\kappa^*$, i.e. κ is purely imaginary. We set conventionally $\kappa = (i\hbar)^{-1}$, where $\hbar$ is a real constant, so

$$[\mathbf{q},\mathbf{p}] \equiv \mathbf{qp} - \mathbf{pq} = i\hbar.$$

We can thus write Eq.(2.6) in the form

$$(x-y)\langle x|\mathbf{p}|y\rangle = i\hbar\langle x|y\rangle = i\hbar\delta(x-y), \tag{2.9}$$

where Eq.(2.2) has been used. A solution of this equation, in the space of tempered distributions, is

$$\langle x|\mathbf{p}|y\rangle = i\hbar\frac{\partial}{\partial y}\delta(x-y) = i\hbar\frac{\partial}{\partial y}\langle x|y\rangle. \tag{2.10}$$

Proof that Eq.(2.10) satisfies Eq.(2.9)

Let $g(x)$ be a differentiable test-function. Then

$$\begin{aligned} & \int dx g(x)(x-y)\frac{\partial}{\partial y}\delta(x-y) \\ = & -\int dx\, g(x)(x-y)\frac{\partial}{\partial x}\delta(x-y) \\ = & \int dx\, \delta(x-y)\frac{\partial}{\partial x}\left[g(x)(x-y)\right] \\ = & \left[g'(x)(x-y)+g(x)\right]_{x=y} = g(y). \end{aligned} \tag{2.11}$$

△

The most general solution of Eq.(2.9), in the space of tempered distributions, is actually

$$\langle x|\mathbf{p}|y\rangle = \left[i\hbar\frac{\partial}{\partial y} + f(y)\right]\delta(x-y), \tag{2.12}$$

where $f(y)$ is an arbitrary real, continuous function (f must be real since p is Hermitian). However, for any x, set

$$|x'\rangle = \exp\left[-\frac{i}{\hbar}\int_{x_0}^{x} d\omega f(\omega)\right]|x\rangle\,.$$

From Eq.(2.2) it is clear that

$$\langle x'|y'\rangle = \delta(x-y),$$

and from Eq.(2.12) it follows that

$$\langle x'|\mathbf{p}|y'\rangle = \exp\left[-\frac{i}{\hbar}\int_{x}^{y} d\omega f(\omega)\right]\left[i\hbar\frac{\partial}{\partial y} + f(y)\right]\delta(x-y)\,.$$

On multiplying this by a test function and integrating, we obtain exactly the same expression as we would have obtained from

$$i\hbar\frac{\partial}{\partial y}\delta(x-y)\,,$$

which means that they are identical distributions. In other words, we have shown that the general solution Eq.(2.12) can be reduced to the special case Eq.(2.10) by a suitable redefinition of the phase of the eigenfunctions $|x\rangle$, and that this redefinition respects the normalization Eq.(2.2). For a free particle, we make this phase choice, but in the presence of interactions it can prove advantageous to make f equal to a physical field.

A physical system is described by a state vector $|\psi\rangle$. This state vector is abstract, in the sense that it is an element of a linear vector space — the same Hilbert space that contains the

eigenfunctions $|x\rangle$ of $\mathbf{q}$. The wave-function of the state *in the configuration representation* is defined to be

$$\psi(x) = \langle x|\psi\rangle\,.$$

The wave-function, unlike $|\psi\rangle$, is simply a (complex) number. The probability of finding the particle in the interval $a \le x \le b$, given that the state vector is $|\psi\rangle$, is postulated to be

$$\int_a^b dx\,|\psi(x)|^2.$$

Consider the space spanned by the eigenvectors $|x\rangle$ of $\mathbf{q}$. That is, consider the set of all vectors that can be written in the form

$$|\phi\rangle = \int dy\,\phi(y)|y\rangle\,, \tag{2.13}$$

where $\phi(y)$ is a (complex) number. Then Eq.(2.2) implies

$$\langle x|\phi\rangle = \int dy\phi(y)\delta(y-x) = \phi(x)\,.$$

Inserting this into Eq.(2.13), we find

$$|\phi\rangle = \int dy|y\rangle\langle y|\phi\rangle\,;$$

and since $|\phi\rangle$ is an arbitrary vector in the space, we may express this in the form

$$\int dy|y\rangle\langle y| = 1\,, \tag{2.14}$$

where the right side here means the **unit operator in the space spanned by the eigenvectors** $|x\rangle$.

Consider the matrix element

$$\begin{aligned}
\langle x|\mathbf{p}|\psi\rangle &= \int dy\,\langle x|\mathbf{p}|y\rangle\langle y|\psi\rangle \\
&= i\hbar\int dy\left(\frac{\partial}{\partial y}\delta(x-y)\right)\langle y|\psi\rangle \\
&= -i\hbar\int dy\,\delta(x-y)\left(\frac{\partial}{\partial y}\langle y|\psi\rangle\right) \\
&= -i\hbar\frac{\partial}{\partial x}\langle x|\psi\rangle \\
&= -i\hbar\frac{\partial}{\partial x}\psi(x)\,,
\end{aligned} \tag{2.15}$$

where Eq.(2.10) has been used.

The result Eq.(2.15) is important: it is sometimes expressed by saying that the operator $\mathbf{p}$ is represented by $-i\hbar\frac{\partial}{\partial x}$. More accurately, we can say that, in the configuration representation, $|\psi\rangle$ is realized by the differentiable function $\psi(x)$, the position operator $\mathbf{q}$ by the real number x, and the momentum operator $\mathbf{p}$ by the *differential operator* $-i\hbar\frac{\partial}{\partial x}$.

It is important to understand that the above realizations of the abstract vectors and operators are specific to the configuration representation. To drive home this point, we will consider next the *momentum representation* as an alternative to the configuration representation. To do this, construct the vector

$$|p\rangle = \frac{1}{2\pi\hbar}\int dy|y\rangle\, \mathrm{e}^{ipy/\hbar}. \tag{2.16}$$

From Eq.(2.10) we see that

$$\begin{aligned}
\langle x|\mathbf{p}|p\rangle &= \frac{1}{2\pi\hbar}\int dy\, \langle x|\mathbf{p}|y\rangle \mathrm{e}^{ipy/\hbar} \\
&= \frac{i\hbar}{2\pi\hbar}\int dy\ \mathrm{e}^{ipy/\hbar}\frac{\partial}{\partial y}\delta(x-y) \\
&= -\frac{i\hbar}{2\pi\hbar}\int dy\, \delta(x-y)\frac{\partial}{\partial y}\,\mathrm{e}^{ipy/\hbar} \\
&= \frac{p}{2\pi\hbar}\int dy\, \langle x|y\rangle\ \ \mathrm{e}^{ipy/\hbar} = p\langle x|p\rangle.
\end{aligned} \tag{2.17}$$

Using Eq.(2.14) and restricting $\mathbf{p}$ to the space spanned by the vectors $|x\rangle$, we have

$$\mathbf{p}|p\rangle = p|p\rangle,$$

i.e. the vector defined in Eq.(2.16) is an eigenvector of the momentum operator $\mathbf{p}$ with eigenvalue p. Clearly

$$\langle x|p\rangle = \frac{1}{2\pi\hbar}\int dy\, \langle x|y\rangle \mathrm{e}^{ipy/\hbar} = \frac{1}{2\pi\hbar}\int dy\, \delta(x-y)\, \mathrm{e}^{ipy/\hbar} = \frac{1}{2\pi\hbar}\mathrm{e}^{ipx/\hbar}\,. \tag{2.18}$$

The state vector $|\psi\rangle$ is realized in the momentum representation by the function

$$\tilde{\psi}(p) = \langle p|\psi\rangle\,,$$

and

$$\begin{aligned}
\langle p|\mathbf{q}|\psi\rangle &= \int dy\, \langle p|\mathbf{q}|y\rangle\langle y|\psi\rangle \\
&= \frac{1}{2\pi\hbar}\int dy\, y\, \mathrm{e}^{-ipy/\hbar}\,\langle y|\psi\rangle \\
&= \frac{i\hbar}{2\pi\hbar}\int dy\left(\frac{\partial}{\partial p}\,\mathrm{e}^{-ipy/\hbar}\right)\langle y|\psi\rangle \\
&= i\hbar\frac{\partial}{\partial p}\int dy\, \langle p|y\rangle\langle y|\psi\rangle \\
&= i\hbar\frac{\partial}{\partial p}\,\tilde{\psi}(p)\,.
\end{aligned} \tag{2.19}$$

In other words, the position operator $\mathbf{q}$ is realized in the momentum representation by the differential operator $i\hbar\frac{\partial}{\partial p}$, the momentum operator $\mathbf{p}$ being realized now by the real number p.

There is accordingly symmetry between the treatment of the configuration and momentum repesentations. In fact, from Eq.(2.18) we see that the relation between the configuration and

momentum representations of a vector is that of Fourier transformation:

$$\begin{aligned}\tilde{\psi}(p) &= \langle p|\psi\rangle \\ &= \int dx\, \langle p|x\rangle\langle x|\psi\rangle \\ &= \frac{1}{2\pi\hbar}\int dx\, \mathrm{e}^{-ipx/\hbar}\psi(x)\,. \end{aligned} \tag{2.20}$$

2.2 The Hamiltonian

Let $|\psi(t)\rangle$ be an arbitrary, normalized state vector of a dynamical system. As a result of the dynamics, this vector will in general change with time. The dynamics implies the existence of an evolution operator, $U(t)$, such that

$$|\psi(t)\rangle = U(t)|\psi(0)\rangle\,.$$

It is supposed that $U(t)$ is a *linear unitary* operator. This is a basic assumption in quantum mechanics: in particular it means that the normalization of a state does not alter in time:

$$\langle\psi(t)|\psi(t)\rangle = \langle\psi(0)|U^\dagger(t)U(t)|\psi(0)\rangle = \langle\psi(0)|\psi(0)\rangle = 1\,.$$

The generator of time translations, H, is closely connected to U. It satisfies

$$H|\psi(t)\rangle = i\hbar\frac{d}{dt}|\psi(t)\rangle \tag{2.21}$$

for any state vector $|\psi(t)\rangle$. It follows that

$$HU(t)|\psi(0)\rangle = i\hbar\dot{U}(t)|\psi(0)\rangle\,,$$

and since this must hold for any state vector $|\psi(0)\rangle$, we conclude that

$$HU(t) = i\hbar\dot{U}(t)\,.$$

Because of the assumed unitarity of the evolution operator, this implies

$$H = i\hbar\dot{U}(t)U^\dagger(t)\,,$$

and hence

$$H^\dagger = -i\hbar U(t)\dot{U}^\dagger(t)\,.$$

However, since $UU^\dagger = 1$ at all times, we see that

$$\dot{U}(t)U^\dagger(t) + U(t)\dot{U}^\dagger(t) = 0\,,$$

so that the time-translation operator is Hermitian:

$$H = H^\dagger\,. \tag{2.22}$$

We shall now show why the operator H, the generator of time-translations, is to be identified as the quantum version of the Hamiltonian. To do this we observe that for any linear operator, $\mathbf{p}$ (such as the momentum for instance), we have

$$\begin{aligned}\langle\psi(t)|\mathbf{p}|\psi(t)\rangle &= \langle\psi(0)|U^\dagger(t)\mathbf{p}U(t)|\psi(0)\rangle \\ &= \langle\psi(0)|\mathbf{p}_H|\psi(0)\rangle\,, \end{aligned} \tag{2.23}$$

where

$$\mathbf{p}_H = U^\dagger(t)\mathbf{p}U(t)$$

is called the *Heisenberg picture* version of the operator $\mathbf{p}$. The Heisenberg picture version of the vector $|\psi(t)\rangle$ is simply $|\psi(0)\rangle$. In the Heisenberg picture, all the time-dependence is carried by the operators, and none by the state vectors. It is in this picture that H can be identified as the Hamiltonian. If $\mathbf{p}$ does not depend on time, then the time-dependence of $\mathbf{p}_H$ is wholly borne by the evolution operators:

$$\dot{\mathbf{p}}_H = \dot{U}^\dagger \mathbf{p} U + U^\dagger \mathbf{p} \dot{U}\,.$$

Since $\dot{U} = -iHU/\hbar$, it follows that $\dot{U}^\dagger = iU^\dagger H/\hbar$, for we have shown that H is Hermitian. Hence

$$\dot{\mathbf{p}}_H = \frac{i}{\hbar}\left\{U^\dagger H \mathbf{p} U - U^\dagger \mathbf{p} H U\right\} = \left[\mathbf{p}_H, H_H\right]/(i\hbar)\,.$$

Here $\mathbf{p}_H$ and H_H are the Heisenberg picture versions of $\mathbf{p}$ and H. Recalling that the commutator divided by $i\hbar$ is the quantum analogue of the classical Poisson bracket, we observe that this equation has exactly the form of Hamilton's equations of motion, *on condition that we identify H as the quantum mechanical Hamiltonian.* In an isolated classical system, the Hamiltonian is time-independent and is equal to the total energy of the system. In quantum theory we suppose also that in an isolated system the Hamiltonian operator H is time-independent, and that its eigenvalues correspond to the possible energies of the system. The fact that H is Hermitian guarantees the reality of these energies.

2.3 Schrödinger and Klein-Gordon Equations

The generalization of the above to the motion of a particle in 3 dimensions is straightforward. Let $\mathbf{q}_i$ and $\mathbf{p}_i$ be respectively the ith Cartesian components of the position and momentum operators. Then the eigenvector $|\vec{x}\,\rangle = |x_1\rangle|x_2\rangle|x_3\rangle$ of $\mathbf{q}_i$, corresponding to the eigenvalue x_i, satisfies

$$\langle\vec{x}\,|\vec{y}\,\rangle = \delta(x_1 - y_1)\delta(x_2 - y_2)\delta(x_3 - y_3) \equiv \delta^3(\vec{x} - \vec{y}\,), \tag{2.24}$$

and the quantization condition is

$$[\mathbf{q}_i, \mathbf{p}_j] = i\hbar\delta_{ij}\,. \tag{2.25}$$

From this it follows in general that

$$\langle\vec{x}\,|\mathbf{p}_i|\vec{y}\,\rangle = \left[i\hbar\frac{\partial}{\partial y_i} + f_i(\vec{x}\,)\right]\delta^3(\vec{x} - \vec{y}\,)\,, \tag{2.26}$$

where $\vec{f}\,(\vec{x}\,)$ is a continuous function. For a free particle we use the phase freedom, as in the one-dimensional case, to remove this function, leaving[1]

$$\langle\vec{x}\,|\mathbf{p}_i|\vec{y}\,\rangle = i\hbar\frac{\partial}{\partial y_i}\delta^3(\vec{x} - \vec{y}\,)\,. \tag{2.27}$$

For an arbitrary state vector $|\phi\rangle$ in the Hilbert space spanned by the eigenfunctions of $\vec{\mathbf{q}}$,

$$|\phi\rangle = \int d^3y |\vec{y}\,\rangle\langle\vec{y}\,|\phi\rangle\,.$$

[1] In the presence of an electromagnetic interaction, however, the phase freedom is used to set $\vec{f}\,(\vec{x}\,) = -\frac{e}{c}\vec{A}\,(\vec{x}\,)$ where $\vec{A}$ is the electromagnetic vector potential.

Hence

$$\begin{aligned}\langle \vec{x}\,|\mathbf{p}_i|\psi(t)\rangle &= \int d^3y\,\langle \vec{x}\,|\mathbf{p}_i|\vec{y}\,\rangle\langle \vec{y}\,|\psi(t)\rangle \\ &= -i\hbar\frac{\partial}{\partial x_i}\langle \vec{x}\,|\psi(t)\rangle \\ &= -i\hbar\frac{\partial}{\partial x_i}\psi(t,\vec{x}\,)\,, \end{aligned} \tag{2.28}$$

where $|\psi(t)\rangle$ is a state vector describing a physical system at time t. Accordingly, we can say that, in the configuration representation, $\mathbf{p}_i$ is represented by the operator $-i\hbar\frac{\partial}{\partial x_i}$:

$$\vec{\mathbf{p}} \longrightarrow -i\hbar\vec{\nabla}\ . \tag{2.29}$$

In an isolated *classical* system, in which the Lagrangian is not an explicit function of the time, the Hamiltonian is equal to the total energy of the system, which is time-independent. In the nonrelativistic mechanics of a particle of mass m, we have

$$H = \frac{\vec{p}\cdot\vec{p}}{2m} + V(\vec{x}\,)\,,$$

where $V(\vec{x}\,)$ is the (conservative) potential.

In making the transition to quantum mechanics, we simply replace the Hamiltonian, the momentum and the position by the corresponding linear operators that have these quantities as their eigenvalues:

$$H = \frac{\mathbf{p}^2}{2m} + V(\vec{\mathbf{q}}\,)\,, \tag{2.30}$$

where $\mathbf{p}^2 = \vec{\mathbf{p}}\cdot\vec{\mathbf{p}}$. If a physical system is described by the state vector $|\psi(t)\rangle$, then this vector will be an eigenvector of the Hamiltonian:

$$i\hbar\frac{d}{dt}|\psi(t)\rangle = H|\psi(t)\rangle = E|\psi(t)\rangle\,, \tag{2.31}$$

where E is the total energy of the system. Hence

$$E\psi(t,\vec{x}\,) = \langle \vec{x}\,|H|\psi(t)\rangle = \langle \vec{x}\,|\frac{\mathbf{p}^2}{2m} + V(\vec{\mathbf{q}}\,)|\psi(t)\rangle\,. \tag{2.32}$$

To evaluate the right side, note that

$$\begin{aligned}\langle \vec{x}\,|\mathbf{p}_i\mathbf{p}_i|\psi(t)\rangle &= \int d^3y\,\langle \vec{x}\,|\mathbf{p}_i|\vec{y}\,\rangle\langle \vec{y}\,|\mathbf{p}_i|\psi(t)\rangle \\ &= \hbar^2\int d^3y\,\left[\frac{\partial}{\partial y_i}\delta^3(\vec{x}-\vec{y}\,)\right]\frac{\partial}{\partial y_i}\psi(t,\vec{y}\,) \\ &= -\hbar^2\nabla^2\psi(t,\vec{x}\,)\,, \end{aligned} \tag{2.33}$$

where use has been made of Eq.(2.27)-Eq.(2.28). Since the eigenvalue of $\mathbf{q}_i$ belonging to $\langle \vec{x}\,|$ is x_i, and the eigenvalue of any power ${\mathbf{q}_i}^n$ is x_i^n, the eigenvalue of $V(\vec{\mathbf{q}}\,)$ is $V(\vec{x}\,)$ (at any rate for a $V(\vec{x}\,)$ that has a Fourier transform). Hence from Eq.(2.32) we find

$$i\hbar\frac{\partial}{\partial t}\psi(t,\vec{x}\,) = E\psi(t,\vec{x}\,) = -\frac{\hbar^2}{2m}\nabla^2\psi(t,\vec{x}\,) + V(\vec{x}\,)\psi(t,\vec{x}\,)\,. \tag{2.34}$$

This is the famous (time-dependent) Schrödinger equation.

Since E is time-independent under the conditions stipulated above, we can solve Eq.(2.31):

$$|\psi(t)\rangle = \exp\left(\frac{-iEt}{\hbar}\right)|\psi(0)\rangle\,.$$

Hence

$$\psi(t,\vec{x}\,) \equiv \langle\vec{x}\,|\psi(t)\rangle = \exp\left(\frac{-iEt}{\hbar}\right)\psi(\vec{x}\,)\,,$$

where

$$\psi(\vec{x}\,) = \psi(0,\vec{x}\,)\,. \tag{2.35}$$

Thus the time dependence can be factored out of Eq.(2.34), yielding

$$E\psi(\vec{x}\,) = \left[-\frac{\hbar^2}{2m}\nabla^2 + V(\vec{x}\,)\right]\psi(\vec{x}\,)\,, \tag{2.36}$$

which is called Schrödinger's time-independent equation.

When the speed of the particle is not small compared to that of light, the mechanics of special relativity must be used; and instead of Eq.(2.30) we have

$$H^2 = \vec{p}\cdot\vec{p}\,c^2 + m^2c^4\,. \tag{2.37}$$

Here we treat a free particle only — a scalar potential cannot be added without ruining relativistic covariance. We will later see how the electromagnetic 4-potential can be incorporated in a relativistically covariant manner. In quantum mechanics, we again replace H and p by the corresponding operators. Sandwiching both sides of Eq.(2.37) between $\langle\vec{x}\,|$ and $|\psi(t)\rangle$ and putting everything on the left, we find

$$\langle\vec{x}\,|H^2 - \mathbf{p}_i\mathbf{p}_i c^2 - m^2c^4|\psi(t)\rangle = 0.$$

By using Eq.(2.27)-Eq.(2.28) and Eq.(2.21), we deduce

$$\left(\partial^2 + \frac{m^2c^2}{\hbar^2}\right)\psi(t,\vec{x}\,) = 0\,, \tag{2.38}$$

where

$$\partial^2 = \frac{1}{c^2}\frac{\partial^2}{\partial t^2} - \nabla^2\,.$$

This is the famous Klein-Gordon equation that replaces the Schrödinger equation in the relativistic domain.

2.4 Dirac Equation

So far we have discussed the equation of motion of a spinless particle. For the electron we need a state vector with two components, $|\psi_i(t)\rangle\,,\ i = 1,2$, in order to account for the two possible eigenvalues of the z-component of the spin ($\pm\hbar/2$). We require each component of $\psi(t,\vec{x}\,) = \langle\vec{x}\,|\psi(t)\rangle$ to satisfy the Klein-Gordon equation, so that relativity will be respected.

However, we might expect the components to be connected in some way, and we follow Dirac in looking for an equation of the form

$$\left[i\gamma^\mu \partial_\mu - \frac{mc}{\hbar}\right] \psi(t,\vec{x}\,) = 0\,. \tag{2.39}$$

Here γ^μ, $\mu = 0,1,2,3$ are four constant matrices, the *Dirac matrices*, that we will determine; and ∂_μ is the relativistic, covariant derivative operator, with components

$$\partial_0 = \frac{1}{c}\frac{\partial}{\partial t} \qquad \partial_1 = \frac{\partial}{\partial x_1} \qquad \partial_2 = \frac{\partial}{\partial x_2} \qquad \partial_3 = \frac{\partial}{\partial x_3}\,.$$

In Eq.(2.39) summation over the repeated index, μ, is implicit, the so-called Einstein convention, and $\psi(t,\vec{x}\,)$ is a 'matrix' with one column that contains its two spin components. The Dirac matrices multiply this column in the standard matrix fashion.

Dirac was motivated to try an equation of the form of Eq.(2.39) by the consideration that the classical Hamilton equations are *linear* in time derivatives, and he wanted to have this feature also in the new quantum theory. However, relativistic covariance requires then linearity in the spatial derivatives too. Dirac's motivation was not sound, since the Klein-Gordon equation, in which the time derivative appears quadratically, describes spinless particles satisfactorily. Moreover, the equations of motion for the quantized electromagnetic field are also quadratic in the time derivative. The best that can be said at this stage is that Eq.(2.39) is a shot in the dark[2]. The ultimate justification for the Dirac equation is experimental and *a posteriori*. A theoretical framework arises in the group-theoretical treatment of the Lorentz group: its covering group, $SL(2,C)$, has spinor representations. This matter lies however beyond the scope of this course.

The great game now is to determine the Dirac matrices in such a way that each component of $\psi(t,\vec{x}\,)$ satisfies the Klein-Gordon equation, which we rewrite in the form

$$\left(g^{\mu\nu}\partial_\mu\partial_\nu + \frac{m^2c^2}{\hbar^2}\right)\psi(t,\vec{x}\,) = 0\,, \tag{2.40}$$

with summation over μ and ν, where $g^{\mu\nu}$ is the metric tensor, defined by

$$g^{00} = 1 \qquad g^{11} = -1 \qquad g^{22} = -1 \qquad g^{33} = -1$$

with all the nondiagonal terms vanishing. Multiply Eq.(2.39) from the left by $[i\gamma^\nu\partial_\nu + mc/\hbar]$ and multiply out the terms:

$$\left[-\gamma^\nu\gamma^\mu\partial_\nu\partial_\mu - i\frac{mc}{\hbar}\gamma^\nu\partial_\nu + i\frac{mc}{\hbar}\gamma^\mu\partial_\mu - \frac{m^2c^2}{\hbar^2}\right]\psi(t,\vec{x}\,) = 0\,. \tag{2.41}$$

If the Dirac matrices satisfy

$$\gamma^\mu\gamma^\nu + \gamma^\nu\gamma^\mu = 2g^{\mu\nu}\,, \tag{2.42}$$

where it is understood that the right side here multiplies a unit matrix, then the following identity between differential operators holds:

$$\begin{aligned}\gamma^\nu\gamma^\mu\partial_\nu\partial_\mu &= \tfrac{1}{2}\left[\gamma^\mu\gamma^\nu + \gamma^\nu\gamma^\mu\right]\partial_\mu\partial_\nu \\ &= g^{\mu\nu}\partial_\mu\partial_\nu\,.\end{aligned} \tag{2.43}$$

[2]See A. Koestler, 'The Sleepwalkers', Hutchinson (1959), for earlier examples in science of genial shots in the dark!

Hence, on condition that the Dirac matrices satisfy the anticommutation relations Eq.(2.42), the Dirac equation is consistent with the Klein-Gordon equation.

What have we gained by this playing with mathematical equations? First of all, it turns out to be impossible to satisfy Eq.(2.42) with four 2×2 matrices!! The Pauli matrices,

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \qquad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \qquad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

satisfy

$$\sigma_i \sigma_j + \sigma_j \sigma_i = 2\delta_{ij} \,,$$

for $i, j = 1, 2, 3$, but there is no fourth matrix that anticommutes with the Pauli matrices. A lesser sleepwalker than Dirac would have given up; but after much worrying Dirac realized that his anticommutation relations could be realized by 4×4 matrices! This means that $\psi(t, \vec{x})$ must be considered to be a column with not two but four components. We expected two components to account for the spin degree of freedom. What is the meaning of the extra two components? Although this is far from obvious, they correspond to the spin states of the *positron*, the antiparticle of the electron, a particle with the same mass and spin as the electron, but with a charge of the same magnitude but *opposite sign*. Playing about with equations, Dirac opened up the world of antimatter!

A much-used representation of the Dirac matrices is

$$\gamma^0 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \equiv \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}$$

$$\gamma^1 = \begin{pmatrix} 0 & \sigma_1 \\ -\sigma_1 & 0 \end{pmatrix} \equiv \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & -1 & 0 & 0 \\ -1 & 0 & 0 & 0 \end{pmatrix}$$

$$\gamma^2 = \begin{pmatrix} 0 & \sigma_2 \\ -\sigma_2 & 0 \end{pmatrix} \equiv \begin{pmatrix} 0 & 0 & 0 & -i \\ 0 & 0 & i & 0 \\ 0 & i & 0 & 0 \\ -i & 0 & 0 & 0 \end{pmatrix}$$

$$\gamma^3 = \begin{pmatrix} 0 & \sigma_3 \\ -\sigma_3 & 0 \end{pmatrix} \equiv \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \\ -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} .$$

We can write this compactly as follows:

$$\begin{aligned} \gamma^0 &= \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \\ \vec{\gamma} &= \begin{pmatrix} 0 & \vec{\sigma} \\ -\vec{\sigma} & 0 \end{pmatrix} . \end{aligned} \tag{2.44}$$

Here 0 and 1 are to be understood respectively as the null and the unit 2×2 matrices.

Note that γ^0 is hermitian and that its square is the unit matrix. The *spacelike* γ's, i.e. γ^1, γ^2, γ^3, are antihermitian, but since they anticommute with γ^0, it is true that for all four Dirac matrices

$$\gamma^0 \gamma^{\mu\dagger} \gamma^0 = \gamma^\mu \,. \tag{2.45}$$

The Dirac equation is often written in the super-compact form

$$[i\gamma\partial - m]\,\psi(x) = 0\,, \tag{2.46}$$

where $\gamma\partial$ simply means the relativistic invariant $\gamma^\mu \partial_\mu$ and x stands for all four dimensions of space-time. To make the equation more beautiful, units of velocity and action have been adjusted in such a way that $c = 1$ and $\hbar = 1$, so that these ugly factors can be omitted. This is common practice in high-energy physics; but in this course we will most of the time retain c and $\hbar$ to avoid possible confusion.

The hermitian conjugate of Eq.(2.46) is $\psi^\dagger(x)[-i\gamma^\dagger \overleftarrow{\partial} - m] = 0$, where $\psi^\dagger(x)$ has been placed to the left: it is a matrix with one row. Multiply by $-\gamma^0$ from the right and use Eq.(2.45):

$$\overline{\psi}(x)\left[i\gamma \overleftarrow{\partial} + m\right] = 0\,, \tag{2.47}$$

where $\overline{\psi}(x) = \psi^\dagger(x)\gamma^0$. This is called the conjugate Dirac equation.

It should be noted that Eq.(2.44) is only one possible representation of the Dirac matrices. If we replace all gamma matrices γ^μ by $\tilde{\gamma}^\mu$, where

$$\tilde{\gamma}^\mu = M\gamma^\mu M^\dagger\,, \tag{2.48}$$

and the four-component wave-function $\psi(x)$ by $\tilde{\psi}(x) = M\psi(x)$, where M is any unitary 4×4 matrix, then all of the equations of this section retain their form. The Dirac matrices are only defined up to a unitary transformation: the important thing is the relation between them.

Appendix A

Integral Orbital Angular Momentum

From the commutation relations that are satisfied by the components of the orbital angular momentum operator, we know that the only possible eigenvalues of L^2 are $\ell(\ell+1)\hbar^2$, where ℓ is a non-negative integer or half-integer, while the only possible eigenvalues of L_3, for a given ℓ, are $\mathsf{m}\hbar$, where m must have one of the values $-\ell, -\ell+1, \ldots, \ell$. In almost all quantum mechanics text books, an inadequate explanation is given for the reason that the half-odd integer values are excluded for the orbital angular momentum. One such inadequate justification is that the spherical harmonics are not one-valued for nonintegral ℓ. However, the lack of one-valuedness for the matrix representations of spin angular momentum is calmly accepted later on, and the 'justification' in the orbital angular momentum case is then conveniently forgotten. The real reason that the half-odd integer values are excluded is that the components of $\vec{L}$ satisfy **more** than merely the Lie algebra corrsesponding to SU(2), as we shall now show.

Proof that m must be an integer

Define

$$Q_\pm = \frac{q_1 \pm p_2}{\sqrt{2}}$$
$$P_\pm = \frac{p_1 \mp q_2}{\sqrt{2}},$$

where q_j and p_j are the position and momentum components of a particle. Then the commutation relations between the q's and the p's imply

$$\begin{aligned} [Q_+, Q_-] &= 0 = [P_+, P_-] \\ [Q_\pm, P_\pm] &= i\hbar \\ [Q_\pm, P_\mp] &= 0, \end{aligned} \tag{A.1}$$

i.e. the canonical pairs (Q_+, P_+) and (Q_-, P_-) are independent of each other. Moreover,

$$\begin{aligned} L_3 &= q_1 p_2 - q_2 p_1 \\ &= \frac{1}{2}(P_+^2 + Q_+^2) - \frac{1}{2}(P_-^2 + Q_-^2). \end{aligned} \tag{A.2}$$

Note that this just looks like the difference between the hamiltonians of two simple harmonic oscillators. We exploit this isomorphism in the following treatment, which we give in full.

Define annihilation and creation operators

$$\begin{aligned} a_\pm &= \frac{Q_\pm + iP_\pm}{\sqrt{2\hbar}} \\ a_\pm^\dagger &= \frac{Q_\pm - iP_\pm}{\sqrt{2\hbar}} \end{aligned} \tag{A.3}$$

so that

$$[a_\pm, a_\pm^\dagger] = 1,$$

with all other commutators vanishing.

Define

$$N_\pm = a_\pm^\dagger a_\pm, \tag{A.4}$$

so that

$$\begin{aligned} [N_\pm, a_\pm] &= -a_\pm \\ [N_\pm, a_\pm^\dagger] &= a_\pm^\dagger. \end{aligned} \tag{A.5}$$

Let $|\nu_\pm\rangle$ be a normalized eigenvector of $N_\pm$ with eigenvalue $\nu_\pm$, i.e.

$$N_\pm|\nu_\pm\rangle = \nu_\pm|\nu_\pm\rangle. \tag{A.6}$$

Then

$$\nu_\pm\langle\nu_\pm|\nu_\pm\rangle = \langle\nu_\pm|a_\pm^\dagger a_\pm|\nu_\pm\rangle$$

or

$$\nu_\pm\||\nu_\pm\rangle\|^2 = \|a_\pm|\nu_\pm\rangle\|^2 \geq 0. \tag{A.7}$$

Hence $\nu_\pm$ are necessarily non-negative real numbers. However, from Eq.(A.5), we find

$$\begin{aligned} N_\pm a_\pm|\nu_\pm\rangle &= (a_\pm N_\pm - a_\pm)|\nu_\pm\rangle \\ &= (\nu_\pm - 1)a_\pm|\nu_\pm\rangle, \end{aligned} \tag{A.8}$$

so that $a_\pm|\nu_\pm\rangle$ is an eigenvector of $N_\pm$ with eigenvalue $\nu_\pm - 1$, unless it vanishes, in which case the above equation is trivially true.

By repeated applications of the lowering operators, $a_\pm$, we can generate smaller and smaller eigenvalues of $N_\pm$, namely $\nu_\pm - 1$, $\nu_\pm - 2$, ..., which would eventually reach negative values, in violation of Eq.(A.7). The only way to avoid the contradiction is if there is an eigenvector of smallest eigenvalue, say $|\nu_{\min\pm}\rangle$, for which

$$a_\pm|\nu_{\min\pm}\rangle = 0; \tag{A.9}$$

because then Eq.(A.8) does not imply that there is an eigenvalue $\nu_{\min\pm} - 1$. However this implies

$$N_\pm|\nu_{\min\pm}\rangle = 0,$$

i.e. $\nu_{\min\pm} = 0$, and so we can replace the above equation by

$$N_\pm|0\rangle = 0, \tag{A.10}$$

where $|0\rangle$ can be taken to be a simultaneous eigenvector of N_+ and N_-, since these operators commute with one another.

From this ground-state $|0\rangle$ we can build up excited states with the help of Eq.(A.5):

$$N_\pm a_\pm^\dagger|0\rangle = (a_\pm^\dagger N_\pm + a_\pm^\dagger)|0\rangle = a_\pm^\dagger|0\rangle$$

so that $a_\pm^\dagger|0\rangle$ is an eigenvector of $N_\pm$ with eigenvalue unity. By repeated applications of the raising operator, $a_\pm^\dagger$, we can build up successively vectors with eigenvalues 2, 3, 4, ... Glancing back at Eq.(A.6), we see that we have established that the eigenvalues $\nu_\pm$ of the operators $N_\pm$ are precisely the non-negative integers.

From Eq.(A.3), Eq.(A.4) and Eq.(A.1) we find that

$$\begin{aligned} N_\pm &= \frac{1}{2\hbar}\{(Q_\pm - iP_\pm)(Q_\pm + iP_\pm)\} \\ &= \frac{1}{2\hbar}\{Q_\pm^2 + P_\pm^2 + i[Q_\pm, P_\pm]\} \\ &= \frac{1}{2\hbar}\{Q_\pm^2 + P_\pm^2 - \hbar\}. \end{aligned} \tag{A.11}$$

Hence Eq.(A.2) yields

$$L_3 = (N_+ - N_-)\hbar\,;$$

and since we have just proved that ν_+ and ν_- are non-negative *integers*, it follows that the eigenvalues of L_3 have the form $m\hbar$, **where m must be an integer!!**. Since m must have one of the values $-l$, $-l+1$, $-l+2$, ..., l, it follows that l must also be an integer. $\triangle$

Appendix B

Dirac Delta Function

In this appendix we will develop the theory of the Dirac delta function, or more properly the delta distribution, along the lines of the distribution theory of L. Schwartz. This abstract approach is not difficult, and it is much more satisfactory than the conventional expositions given in many quantum mechanics books.

A *functional* is a mapping from a function space onto $\mathcal{C}$, the set of complex numbers. For example, the integral

$$\int_{-\infty}^{\infty} dx\, g(x) = G$$

is a representative of the mapping of a space of functions, g, onto their integrals, taken over the real line. Of course, this only makes sense if g is integrable. A space that we will use below is that of all infinitely differentiable functions of compact support; that is, all functions, $g(x)$, that can be differentiated an arbitrary number of times, and are such that, for each $g(x)$, there is an $x_{\min}$ and an $x_{\max}$, such that

$$g(x) = 0 \qquad \text{unless} \qquad x_{\min} < x < x_{\max}\,.$$

The support $[x_{\min}, x_{\max}]$ will not be the same for each $g(x)$, and in general, for any x there will be some g's that do not vanish there. For this space, the Schwartz space, $\mathcal{S}$, the above integral is well-defined, and so the functional is also defined.

The distribution δ is defined, with respect to the space $\mathcal{S}$, to be the functional that maps any function $g(x) \in \mathcal{S}$ onto the number $g(0)$, i.e. onto its value at $x = 0$. That is all there is to it! Here $\mathcal{S}$ is an example of a space of test functions: one is free to choose other spaces, for example the space of all functions that are merely continuous, on which the definition of the mapping

$$\delta : \qquad g(x) \longrightarrow g(0) \tag{B.1}$$

still makes good sense. However, a distribution like δ is defined *in terms of a space of test functions*. Change that space and you change the distribution!

The mapping (B.1) is usually expressed

$$\int_{-\infty}^{\infty} dx\, g(x)\delta(x) = g(0)\,, \tag{B.2}$$

where $\delta(x)$ is written as if it were a function with argument x: but this is emphatically not the case! In order for an integrable function to be able to 'sample' the test-function, $g(x)$, at just one point, $x = 0$, it would have to vanish at all points $x \neq 0$, but be such that

$$\int_{x_{\min}}^{x_{\max}} dx\, \delta(x) = 1$$

for all $x_{\min} < 0$ and $x_{\max} > 0$. Intuitively, if $\delta(x)$ were a function, it would have to be infinitely sharply peaked at $x = 0$. There are no such functions! That, however, is no problem: Eq.(B.2) is merely a conventional way of writing Eq.(B.1), no more and certainly no less.

Making the change of variables

$$x \longrightarrow x - y \qquad \hat{g}(x) = g(x - y) \qquad \hat{g}(y) = g(0)$$

in Eq.(B.2), we obtain

$$\int_{-\infty}^{\infty} dx\, \hat{g}(x)\delta(x - y) = \hat{g}(y)\,. \tag{B.3}$$

Since $g(x) \in \mathcal{S} \Rightarrow \hat{g}(x) \in \mathcal{S}$, we can effectively drop the caret in Eq.(B.3): the space of all g's is the same as the space of all $\hat{g}$'s, and $\delta(x - y)$ is the distribution that maps a function $g(x)$ onto the value $g(y)$.

Note that $\delta(x)$ is even. We can prove that by making the change of variables

$$x \longrightarrow -x \qquad g(-x) = \hat{g}(x) \qquad g(0) = \hat{g}(0)$$

in Eq.(B.2). This gives

$$\int_{-\infty}^{\infty} dx\, \hat{g}(x)\delta(-x) = \hat{g}(0)\,. \tag{B.4}$$

The distribution $\delta(-x)$ yields exactly the same result as $\delta(x)$ itself; since a distribution is defined as a mapping, and the mappings are the same, it follows that

$$\delta(-x) = \delta(x)\,.$$

We can then also write

$$\int_{-\infty}^{\infty} dx\, \hat{g}(x)\delta(y - x) = \hat{g}(y)\,. \tag{B.5}$$

In order to obtain a useful representation for the delta function, we recall that the complex Fourier transform of $g(x)$ is

$$\overline{g}(k) = \int_{-\infty}^{\infty} dx\ \mathrm{e}^{ikx} g(x)\,, \tag{B.6}$$

the inverse being

$$g(y) = \frac{1}{2\pi} \int_{-\infty}^{\infty} dk\ \mathrm{e}^{-iky} \overline{g}(k)\,. \tag{B.7}$$

These formulae are certainly valid if $g(x) \in \mathcal{S}$, indeed, it is enough for $g(x)$ to be a piecewise continuous function, according to the theory of Fourier analysis. Substituting the first equation above into the second, we obtain

$$\begin{aligned} g(y) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} dk\ \mathrm{e}^{-iky}\int_{-\infty}^{\infty} dx\ \mathrm{e}^{ikx}g(x) \\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty} dx\, g(x)\int_{-\infty}^{\infty} dk\ \mathrm{e}^{ik(x-y)}\,. \end{aligned} \tag{B.8}$$

On comparing this with Eq.(B.3), we see that

$$\delta(x-y) = \frac{1}{2\pi}\int_{-\infty}^{\infty} dk\ \mathrm{e}^{ik(x-y)}\,. \tag{B.9}$$

It should be clear that this integral does not make sense as an integral of a function, yielding another function: it has to be understood *in the sense of distribution theory*, that is, it is merely a shorthand way of expressing Eq.(B.6) and Eq.(B.7). Of course, an intuitive, gut-feeling interpretation of Eq.(B.9) *is* possible: it says that, unless $x = y$, the integral (B.9) *oscillates itself to death!* When $x = y$ the integral gives infinity, so what could it be but the Dirac delta function!?

The interchange in integration orders that we perpetrated in Eq.(B.8) is strictly a formality: the golden rule in distribution theory is that formulae are only to be understood in the usual sense as functions *after multiplication by a test function, followed by integration.*

On taking the derivative of Eq.(B.7), we find

$$g'(y) = -\frac{i}{2\pi}\int_{-\infty}^{\infty} dk\, k\, \mathrm{e}^{-iky}\overline{g}(k)\,; \tag{B.10}$$

and hence

$$\begin{aligned} -g'(0) &= \frac{i}{2\pi}\int_{-\infty}^{\infty} dk\, k\overline{g}(k) \\ &= \frac{i}{2\pi}\int_{-\infty}^{\infty} dk\, k\int_{-\infty}^{\infty} dx\ \mathrm{e}^{ikx}g(x) \\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty} dx\, g(x)\int_{-\infty}^{\infty} dk\,\frac{\partial}{\partial x}\,\mathrm{e}^{ikx} \\ &= \int_{-\infty}^{\infty} dx\, g(x)\delta'(x) \end{aligned} \tag{B.11}$$

where

$$\delta'(x) = \frac{1}{2\pi}\frac{d}{dx}\int_{-\infty}^{\infty} dk\ \mathrm{e}^{ikx} = \frac{d}{dx}\delta(x)\,.$$

In other words, the distribution δ' is the following mapping:

$$\delta' : \qquad g(x) \longrightarrow -g'(0)\,. \tag{B.12}$$

Higher derivatives of the delta function can be similarly defined.

The step-function is

$$\theta(x) = \int_{-\infty}^{x} dy\, \delta(y)\,.$$

As a (discontinuous) function, it is clear that, for $x < 0$, $\theta(x) = 0$, and for $x > 0$, $\theta(x) = 1$, whereas it is not defined — *as a function* — for $x = 0$. Considering the step function rather as a distribution, we have

$$\begin{aligned}\int_{-\infty}^{\infty} dx\, g(x)\theta(x) &= \int_{-\infty}^{\infty} dy\, \delta(y) \int_{y}^{\infty} dx\, g(x) \\ &= \int_{0}^{\infty} dx\, g(x)\,.\end{aligned}$$

For any test-function with support in $(-\infty, 0)$, this yields zero. The derivative of the step-function is the delta-function:

$$\frac{d}{dx}\theta(x) = \delta(x)\,.$$

That this formal manipulation yields a correct result can be checked by playing with test functions:

$$\begin{aligned}\int_{-\infty}^{\infty} dx\, g(x)\frac{d}{dx}\theta(x) &= -\int_{-\infty}^{\infty} dx\, \theta(x)\frac{d}{dx}g(x) \\ &= -\int_{0}^{\infty} dx\, \frac{d}{dx}g(x) \\ &= g(0)\,,\end{aligned}$$

which effects the required proof. Note that, in the partial integration above, the boundary terms have been left out, since the test-functions, $g(x) \in \mathcal{S}$, being of compact support, vanish at infinity. The same proof applies of course for any space of differentiable test-functions that tend to zero as $x \longrightarrow \pm\infty$.

In many elementary treatments, the delta function is introduced as a 'limit', such as

$$\delta(x) = \lim_{\varepsilon \to 0} \frac{1}{\sqrt{\pi\varepsilon}} \exp\left(-\frac{x^2}{\varepsilon}\right)\,. \tag{B.13}$$

This of course does not exist as a limit in the usual sense: we are told to use it only 'under an integral'. In the language of test functions, we write

$$\begin{aligned}&\lim_{\varepsilon \to 0} \frac{1}{\sqrt{\pi\varepsilon}} \int_{-\infty}^{\infty} dx\, g(x) \exp\left(-\frac{x^2}{\varepsilon}\right) \\ =\ &\lim_{\varepsilon \to 0} \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} dy\, g(y\sqrt{\varepsilon}) \exp\left(-y^2\right) \\ =\ &\frac{g(0)}{\sqrt{\pi}} \int_{-\infty}^{\infty} dy \exp\left(-y^2\right) \\ =\ &g(0)\,.\end{aligned}$$

Notice that taking the limit $\varepsilon \to 0$ in $g(y\sqrt{\varepsilon})$ to get $g(0)$ is perfectly legitimate, since the test-function has compact support.

Two other representations for the delta function are

$$\delta(x) = \lim_{\varepsilon \to 0} \frac{1}{\pi} \frac{\varepsilon}{x^2 + \varepsilon^2} \tag{B.14}$$

and

$$\delta(x) = \lim_{N \to \infty} \frac{1}{\pi} \frac{\sin Nx}{x} \tag{B.15}$$

as can be easily checked by means of test function manipulations.

Appendix C

Hilbert Space

A linear vector space, $\mathcal{H}$, is a set of elements, ψ, ϕ, ..., called vectors, such that if $\psi \in \mathcal{H}$ and $\phi \in \mathcal{H}$, then for any complex numbers c and d, $c\psi + d\phi \in \mathcal{H}$. Note that, by choosing $\phi = \psi$ and $c = -d$, we see that the null vector, usually just written 0, necessarily belongs to $\mathcal{H}$.

A scalar product of two vectors, $\mathcal{S}(\psi, \phi)$, is a complex number that satisfies the following conditions:

1. $[\mathcal{S}(\psi, \phi)]^* = \mathcal{S}(\phi, \psi)$
2. $\mathcal{S}(\psi, c\phi + d\chi) = c\mathcal{S}(\psi, \phi) + d\mathcal{S}(\psi, \chi)$
3. (a) $\mathcal{S}(\phi, \phi) > 0$ if $\phi \neq 0$
   (b) $\mathcal{S}(\phi, \phi) = 0$ if $\phi = 0$.

An example of a linear vector space is the set of all $1 \times n$ dimensional matrices of the form

$$\phi = \begin{pmatrix} \phi_1 \\ \phi_2 \\ \phi_3 \\ \cdots \\ \phi_n \end{pmatrix} . \tag{C.1}$$

In matrix algebra nomenclature, such column matrices are indeed usually called vectors. A scalar product on this space is

$$\mathcal{S}(\psi, \phi) = \psi^\dagger \phi = \begin{pmatrix} \psi_1^* & \psi_2^* & \psi_3^* & \cdots & \psi_n^* \end{pmatrix} \begin{pmatrix} \phi_1 \\ \phi_2 \\ \phi_3 \\ \cdots \\ \phi_n \end{pmatrix} = \psi_i^* \phi_i \tag{C.2}$$

with implicit summation over i.

A linear functional $\mathcal{F}$ on a vector space is a mapping from the space onto the complex numbers: that is, for any $\phi_i \in \mathcal{H}$, we have $\mathcal{F}(\phi_i) \in \mathcal{C}$, where $\mathcal{C}$ is the set of all complex numbers, and moreover

$$\mathcal{F}(c\phi_1 + d\phi_2) = c\mathcal{F}(\phi_1) + d\mathcal{F}(\phi_2)\,. \tag{C.3}$$

An example of such a linear functional can be constructed, given a fixed vector, ψ, as in Eq.(C.2). Let us rewrite it

$$\mathcal{F}_\psi(\phi) = \begin{pmatrix} \psi_1^* & \psi_2^* & \psi_3^* & \dots & \psi_n^* \end{pmatrix} \begin{pmatrix} \phi_1 \\ \phi_2 \\ \phi_3 \\ \dots \\ \phi_n \end{pmatrix}$$

The point of putting the subscript ψ on $\mathcal{F}_\psi(\phi)$ is to stress the idea that ψ is to be thought of as a fixed parameter that specifies the functional $\mathcal{F}_\psi$, whilst ϕ is a variable that ranges over the whole vector space. It is obvious that this functional is linear, i.e. it satisfies Eq.(C.3). For every vector $\psi \in \mathcal{H}$ there is a corresponding linear functional $\mathcal{F}_\psi$ on $\mathcal{H}$. The converse is also true: for every linear functional $\mathcal{F}_\psi$, there exists a vector $\psi \in \mathcal{H}$, such that $\mathcal{S}(\psi, \phi) = \mathcal{F}_\psi(\phi)$ — this is called the Riesz Theorem.

Proof of the Riesz Theorem

Given a linear functional $\mathcal{F}_\psi$ and an orthonormal basis, e_n, for which $\mathcal{S}(e_m, e_n) = \delta_{mn}$, construct the vector

$$\psi = \sum_n [\mathcal{F}_\psi(e_n)]^* e_n\,.$$

This does the job, for

$$\psi^\dagger = \sum_n \mathcal{F}_\psi(e_n) e_n^\dagger\,,$$

so

$$\begin{aligned} \mathcal{S}(\psi, \phi) &= \sum_n \mathcal{F}_\psi(e_n)\mathcal{S}(e_n, \phi) \\ &= \sum_n \mathcal{F}_\psi(\mathcal{S}(e_n, \phi)e_n)\,. \end{aligned}$$

Now since $\{e_m\}$ is a basis, ϕ can be expanded on it, say

$$\phi = \sum_m a_m e_m\,,$$

so

$$\begin{aligned} \mathcal{S}(\psi, \phi) &= \sum_n \mathcal{F}_\psi\left(\mathcal{S}\left(e_n, \sum_m a_m e_m\right) e_n\right) \\ &= \sum_m \sum_n a_m \mathcal{F}_\psi(\mathcal{S}(e_n, e_m)e_n) \\ &= \sum_m \sum_n a_m \delta_{mn} \mathcal{F}_\psi(e_n) \\ &= \sum_n a_n \mathcal{F}_\psi(e_n) = \mathcal{F}_\psi(\sum_n a_n e_n) = \mathcal{F}_\psi(\phi)\,. \end{aligned}$$

$\triangle$

Dirac introduced a way of distinguishing notationally between column vectors like Eq.(C.1), which he called *kets* and wrote $|\phi\rangle$, and their hermitian conjugates, which are row vectors, called *bras* and written $\langle\psi|$. The scalar product Eq.(C.2) becomes the bra-c-ket

$$\mathcal{S}(\psi,\phi) = \langle\psi|\phi\rangle\,.$$

The kets $|e_n\rangle$ span the Hilbert space, while their hermitian conjugates, the bras $\langle e_n|$, span the conjugate space. By virtue of the Riesz theorem, the bra $\langle\psi|$ can be identified with the linear operator $\mathcal{F}_\psi$ that we introduced above. Indeed, we may write

$$\mathcal{F}_\psi(|\phi\rangle) = \langle\psi|\phi\rangle\,,$$

or more briefly $\mathcal{F}_\psi = \langle\psi|$. The *dual* of $\mathcal{H}$ is the space of all linear functionals on $\mathcal{H}$, which, because of the Riesz theorem, is isomorphic to $\mathcal{H}$ itself. Indeed for every $\mathcal{F}_\psi \equiv \langle\psi|$ belonging to the dual space, there is a corresponding ket in $\mathcal{H}$ that satisfies

$$\langle e_n|\psi\rangle = [\langle\psi|e_n\rangle]^* = [\mathcal{F}_\psi(|e_n\rangle)]^*$$

A linear vector space, equipped as above with a scalar product, is called a *Hilbert* space if it is complete. A complete space is one that contains all of its limit points. More technically, a Cauchy sequence of vectors, $|\phi_i\rangle$, is a sequence for which, for any $\epsilon > 0$,

$$\exists p : \parallel \phi_i - |\phi_j\rangle \parallel < \epsilon,\ \forall i > p \text{ and } j > p\,, \tag{C.4}$$

where the norm here is defined by

$$\parallel \phi \parallel = \sqrt{\langle\phi|\phi\rangle}\,.$$

If all Cauchy sequences in a space $\mathcal{H}$ are convergent, then there exists for any sequence satisfying Eq.(C.4) a vector $|\phi\rangle \in \mathcal{H}$ such that

$$\exists p : \parallel \phi_i - \phi \parallel < \epsilon,\ \forall i > p\,.$$

The space is then complete and as such is a true Hilbert space. This is automatically the case for the *finite dimensional* spaces that we have been visualizing so far, as in Eq.(C.1); but there is no need to limit our considerations to a finite number of dimensions — indeed it is essential in quantum mechanics often to work with an infinite number of dimensions. If such an infinite-dimensional normed space is not complete it it is called a prehilbert space, but it may be converted into a genuine Hilbert space by adjoining all of its limit points. Such topics, including the rigorous treatment of unbounded operators on Hilbert space, lie beyond the scope of this course.